

UNCLASSIFIED

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

August 1998

19981110091

DITP: A FLEXIBLE MINIATURIZED SENSOR FUSION PROCESSOR FOR NEXT GENERATION INTERCEPTOR SEEKERS AND SURVEILLANCE SENSORS

Dr. Richard W. Linderman
Air Force Research Laboratory (AFRL)
Information Directorate (IFTC)
26 Electronics Pkwy, Bldg. 106
Rome, NY 13441-4514

James C. Lyke
Air Force Research Laboratory (AFRL)
Space Vehicles Directorate (VSSE)
3550 Aberdeen Ave, SE, Kirtland AFB
Albuquerque, NM 87117-5776

Yasuhiro Kinashi
Nichols Research Corporation
2101 Wilson Blvd., Suite 1000
Arlington, VA 22201-3062

ABSTRACT

This paper describes an advanced processor design for the future BMD interceptor seeker and surveillance sensor technology being developed under the sponsorship of the BMDO Technology Office. It offers a unique synergistic solution; efficiently packaging an Embedded High Performance Computer (EHPC) with adaptive, reconfigurable components and mixed analog, digital, microwave and power circuits. While the EHPC provides high performance, programmable, power efficient floating point computations, the adaptive, reconfigurable logic brings flexibility to front end sensor interfaces and tremendous throughput on sensor preprocessing. Advanced packaging allows tremendous internal data bandwidth with 1000 interconnects per layer to accommodate dynamic data/message passing interfaces in real time; also LEGO™-like segments that can be accumulated into a highly dense three-dimensional processor system. This gives the flexibility to combine different types of IC components and MCMs from various vendors within a single assembly. This robust-mixed signal processor system is referred to collectively as the Sensor and Fusion Engine (SAFE). The SAFE EHPC will quadruple the power efficiency of current processors by achieving over 200 million floating point operations per second per watt (MFLOPS/Watt). With a packaging thermal capacity of 500 watts, up to 100 GFLOPS segments deliver an order of magnitude improvement in system density. Simply put, the DITP application of SAFE will constitute the densest three-dimensional system assembly ever attempted.

INTRODUCTION

The Surveillance and Interceptor Technology Directorate in the Ballistic Missile Defense

Organization (BMDO/TOS) has continued to develop advanced technologies needed to counter evolving ballistic missile threats from growing numbers of developing countries. Meeting the demands for next generation interceptor seeker and surveillance sensor electronics will require advanced design and fabrication techniques that not only provide extraordinary miniaturization, but also allow significant performance enhancements to be incorporated. For example, the Air Force's space processing requirements for near term is 200 megabits/sec of sensor data bandwidth, with an ultimate goal of 5.0 gigabits/sec. Further, BMDO has identified an advanced interceptor seeker signal processor need of 20 GFLOPS (billions of floating point operations per second).

DITP Processor System Overview

Challenges for high throughput processors requiring low size, weight, and power (SWAP) are being addressed by the Discriminating Interceptor Technology Program (DITP), a key advanced sensor technology initiative under BMDO/TOS. Its goal is to design and develop both advanced passive and active imaging sensors and a sensor fusion processor, then integrate them into a miniaturized seeker package [1]. For the DITP sensor fusion processor subsystem, a robust mixed-signal processor is designed for 16 gigabits/sec data throughput with processing capacity of 14 to 20 GFLOPS, contained in a miniaturized system with dimensions approximately 2.0 by 2.0 inches on a side by 4.0 inches tall. It combines an advanced embedded high performance computing architecture with flexible high bandwidth data crossbar and power efficient programmable and malleable processors. Major elements include the Wafer Scale Signal Processor (WSSP) being developed by Air Force Research Laboratory (AFRL), Rome, NY; the DARPA-developed Myrinet crossbar interface;

the Malleable Signal Processor (MSP) for a reconfigurable data flow and front end algorithms; and three-dimensional (3-D) heterogeneous multi-chip module (MCM) packaging and system integration from the Highly Integrated Packaging and Processing (HIPP) program at AFRL, Kirtland AFB, NM.

The DITP processor system is designed for flexible integration with a variety of focal plane array (FPA) design configurations. This is accomplished with two sensor adaption segments, one for each FPA, and an MSP crossbar segments. The output of the Ladar signal processor is interfaced to a separate MSP block segment mated to the Myrinet crossbar network segment. The MSP's reconfigurable processor provides the flexibility to adapt the hardware configuration to the needs of command/control structures, data-flow structures, and associated algorithms. The processor gives system designers similar flexibility that software-based simulations provide, but with speed that supports the real-time needs for direct fielding. Therefore, applications can be rapidly synthesized and "injected" into embedded systems. The conventional alternative design approach is fabricating a custom application specific integrated circuit (ASIC), which requires significant time and fabrication expense, "de-integration" of target hardware for component replacement, and re-integration.

The use of reconfigurable processing serves an important role in the general flexibility of an embedded system to deal with change. In the case of a hybrid system, such as DITP, where a reconfigurable processing system is interfaced to a powerful multiprocessing system, WSSP, with a scaleable but fixed architecture, the trade-off can take the form of juggling which operations are performed where. Processor designed to handle high volumes of floating point are not always optimally utilized in "pixel-smashing" operations, and the use of flexible reconfigurable processors, like the MSP, can allow the mapping of many different heuristics without resorting to rebuilding hardware. Alternatively, if the scaleable EHPC has sufficient capacity for an algorithm, it is much faster to reprogram its software, than reconfiguring hardware. The strengths are complementary and work to reduce overall system integration risk.

The heart of the sensor fusion processor subsystem is the EHPC comprised of WSSP segments with a multiple-instruction-multiple-data (MIMD), floating-point Digital Signal Processor

(DSP) architecture. The first-generation WSSP chip-set has been successfully demonstrated at the AFRL, Rome Research Site with power efficiency exceeding 100 MFLOPS/watt (million floating point operations per second per watt) and up to 270 MFLOPS/watt at reduced voltages. This processor performance is at least four times improvement over the state of the art in power density (i.e., MFLOPS/watt), which is a critical DSP performance metric in many embedded applications, including missile tracking, satellite, and advanced airborne systems.

Other Potential Applications

Although the current DITP development effort concentrates on spaceborne applications, which emphasize obtaining miniaturized packaging and power efficiency while increasing throughput and bandwidth, this processor technology is just as suitable for airborne or ground base sensor processing applications.

A separate development for the Air Force is being pursued for potential signal processor enhancements for the Joint Surveillance Targeting and Attack Radar System (JSTARS) and Airborne Warning and Control Systems (AWACS; performing complex matrix operations for adaptive beam-forming to cancel the effects of interfering clutter and jammers. The same power efficiency applies to these and ground-based radar (or sonar) systems to maximize the data bandwidth capacity and the processor throughput (to greater than TeraFLOPS). This translates directly to a cost saving for the production, as well as operations and maintenance costs; hence an overall saving in the total life cycle cost.

Concurrently, the AFRL's Information and the Space Vehicle Directorates are collaborating on a study to explore methods to radiation-harden the processor for harsh space applications in which low power dissipation and miniaturized packaging is of critical concern.

The succeeding sections provide a more detailed description of the three major technology elements that collectively make up the DITP's advanced sensor fusion processor technology. Respectively, they are WSSP, MSP and HIPP.

WAFER SCALE SIGNAL PROCESSOR

The WSSP is a fully programmable, high performance signal processor, designed for high sustained floating point performance for a given amount of power (measured in MFLOPS/Watt),

fault tolerance, and low size, weight and power. The simple computing elements of the WSSP are formed by attaching commercial synchronous static random access memories (SSRAMs) to a low-power processor using a MCM technology supporting area interconnect, such as hybrid wafer scale integration [2]. The WSSP computing element requires approximately 0.7 square inches of MCM area with a 0.5- μ m CMOS processor and 2.0 megabit commercial SRAM parts. Multiple MCMs are interconnected with commercial PCI [3] busses and Myrinet crossbar switches to form a hierarchical multiprocessor environment.

WSSP Processing Element

A WSSP processing element is comprised of a dual processor chip (see photomicrograph in Figure 1) capable of eight single precision floating point operations per clock cycle and two variable depth banks of commercial synchronous SRAMs. Each of the processors has a 72 bit (64 bits plus byte parity) bus to memory. The dual processor chip also includes a shared interface to an external 72-bit I/O bus that supports direct memory access to and from the element memory banks. The two processors, denoted A and B, can each access either memory bank as shown in Figure 2 using the most significant bit of the 32 bit address to identify the memory bank. The memory banks can be up to four, 72 bit wide layers deep. Each layer forms a contiguous portion of the address space.

Element Memory

Using, for example, commercially available 2 megabit synchronous SRAMs organized 64K x 36-bit word, each layer is comprised of two SRAM chips for a minimal memory of 64K x 72-bit words per processor. Alternately, 128K x 18 SRAMs can be used to double the depth of each layer at the expense of nearly double the memory power. This configuration yields a maximum of 512K words per processor using sixteen SRAMs.

The most important feature of the element architecture from a user's viewpoint is the return to flat, SRAM memory. Cache with dynamic random access memory (DRAM) approaches typically supporting RISC (reduced instruction set computing) microprocessors complicate the task of sustaining high bandwidth from the memory to the processor. This has made it difficult for regular users to obtain a high percentage of the advertised peak performance of these processors. Experts can apply tricks such as manual cache management and assembly language optimized functions to achieve over 50% of peak on selected applications, but regular users typically find at most 15 to 25% of peak sustained.

Predictable SRAM memory timing leads to predictable instruction set timing that provides a much needed "rapid prototyping" assist to signal processing system developers confronted with real-time deadlines which must be guaranteed under worst case conditions.

The unpredictability of cache misses and DRAM page opening penalties on highly pipelined RISC microprocessors significantly complicates worst case performance prediction. Current techniques such as speculative execution of instructions yield better data processors but further complicate the implementation of signal processing applications [4].

While software would be easiest to develop for a single high performance processor, the latency and throughput requirements of real-time applications often demand parallel processing approaches. Given a software composer's need to orchestrate activities across parallel processors, the predictable timing resulting from the flat, SRAM primary memory helps plan and achieve synchronization across the players. Since the processing elements require less than one square inch on one layer of an MCM, hundreds can be envisioned in multi-MCM ensembles using message passing and distributed shared memory. Predictable timing assists our ability to reason about, and build tools which accurately model, the behavior of these complex systems.

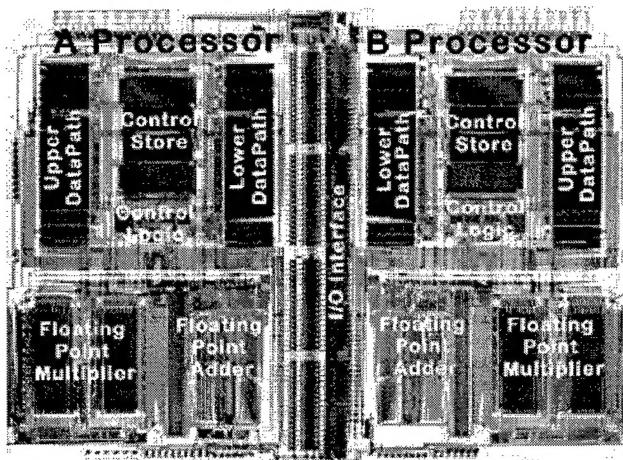


Figure 1. WSSP Processor

Dual Processor Chip Architecture

Relieving the need for on-chip caches frees up more silicon area for the floating point units. As seen in Figure 1, approximately 30% of the silicon area is devoted to the floating point units, the rest can be viewed as overhead if the metric relates to floating point operations per second. Each processor has a floating point arithmetic logic unit (ALU) and a floating point multiplier, each of which can perform either two IEEE single precision operations per clock cycle or one IEEE double precision operation per clock cycle. Supporting both single and double precision creates an opportunity to reach beyond solely signal processing applications to address some traditional supercomputer applications (e.g. computational fluid dynamics) that require 64-bit precision. In addition, some important signal processing applications, such as precision trackers, require 64-bit precision. The overall peak computation rate is eight single precision or four double precision operations per clock cycle. Clock rates are expected to be 50 MHz for 0.5- μ m CMOS, providing peak performance of 400 MFLOPS per chip at an estimated 3.0 watts per processing element (processor and memories).

A Very Long Instruction Word (VLIW) in a 128 bit wide format controls the resources of each processor. Each processor has a dual-3-bus architecture, with upper and lower 32 bit datapaths, two floating point units, two integer units, incrementors and addressing hardware that enable the processor to sustain eight FLOPS/clock cycle on common signal processing operations such as complex dot products. The VLIW instruction store is a combination of RAM and ROM. This writable control store, combined with dual datapaths and computational resources, allows new instructions to be generated dynamically at the assembly level.

Upon the VLIW control structure, a superscalar-vector assembly language instruction set has been written into the ROM control store. The scalar portion of this assembly language instruction set resembles a typical RISC load/store instruction set. To this base are added some superscalar instructions that use the dual ALUs for doubled performance. However, it is the addition of vector instructions, such as Fast Fourier Transform (FFT), dot products, and vector add/subtract/multiply that allow the processors to fully utilize the on-chip computing resources. These instructions typically require clock cycles proportional to vector length. During this time, the full memory bandwidth is available for data

movement since, unlike a RISC architecture, instructions need not be fetched.

The logic throughout the dual processor chip avoids dynamic logic to allow the processor clocks to be disabled for indefinite periods of time without losing state information. This facilitates both testing and element input/output as described below. Each processing element can operate in any of three modes:

1. Independent (Standalone) mode,
2. Coprocessor mode, or
3. Watchdog mode.

In standalone mode, each processor A and B operate completely independently on separate programs, with their own private memory bank as shown in Figure 2. Usually memory bank A is dedicated to processor A and memory bank B to processor B, but not necessarily. When in coprocessor mode, processors A and B work in a client-server relationship to finish a single assignment, such as a long FFT or matrix factorization, with lower latency. In this case, they communicate and share data and results through their ability to access both memory banks and inhibit the other

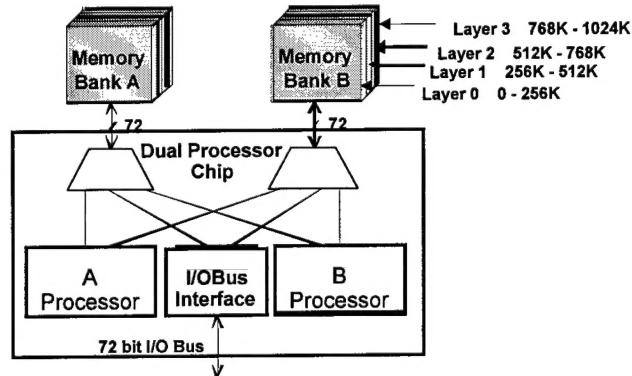


Figure 2. Processor Element Architecture

processor's clocks to guarantee mutually exclusive memory access. In watchdog mode, also known as a self-checking pair configuration, both processors perform exactly the same calculations with lock-step timing, and compare results each clock cycle, to detect errors. If the watchdog processor detects a discrepancy, it will interrupt the "active" processor and halt.

Processing elements can alter modes of operation as the applications are executed. This offers unique possibilities to move from independent to coprocessor mode if latency becomes critical. Alternatively, the processing element can transition to watchdog mode if a critical section of code

is to be executed where concurrent error detection is desired.

Inter-element Communication

The processing elements within the MCM communicate with one another over a multiprocessor bus called the IOBUS (as shown in Figure 3).

The IOBUS is a derivative of the FutureBus+ standard [5] and the PCI local bus [3]. It is 64 bits wide with byte parity checking, and it operates at the speed of the processors. The IOBUS may transfer one 64-bit word per clock cycle. Therefore, at 50 MHz, the bus achieves a peak messaging throughput of 400 megabytes per second.

The FutureBus+ low level messaging format is used by the IOBUS, however, at the signaling level, the IOBUS is synchronous, more like the PCI bus.

The IOBUS does not directly support read requests at the bus level (except for the registers contained in the IOBUS interface itself), these are split into a message requesting information and a write message returning the requested information. Hence, the IOBUS targets distributed shared memory with explicit message passing using MPI (message passing interface).

Real-time systems often require fast responses to external events and changes in internal conditions. Since the IOBUS is the only means by which processors on different elements communicate, when a processor must quickly send a message to an element in the MCM, preemption of lower priority messages is desired.

The IOBUS supports three levels of messaging priority with preemption. The highest priority level is reserved for any external messages coming into the MCM. This external bus is considered a more valuable resource shared by several MCMs. The next level of priority is given to high priority messages originating from within the MCM. The remaining messages are the lowest priority.

A high priority message can preempt a lower priority message within three cycles. Once the high priority transaction is complete, possession of the IOBUS is arbitrated. When the preempted element is again granted the IOBUS, the preempted message will resume where it left off; it does not have to start over.

Transactions on the IOBUS take precedence over the computation on the processors. This is done because the IOBUS is a MCM-wide resource shared among many processors. A stall on the IOBUS could, in turn, stall many of the processors on the MCM; therefore, the IOBUS is given priority. This is accomplished without delay on the IOBUS by inhibiting the clocks of the processors on the receiving element.

In addition to maximizing the effective utilization of the IOBUS, giving the IOBUS priority over the processors yields predictable timing of messages within the MCM. When a processing element is involved in an IOBUS transaction, the clocks for its processors are immediately inhibited. Because the processors are fully static, they may be held in this state indefinitely. With their

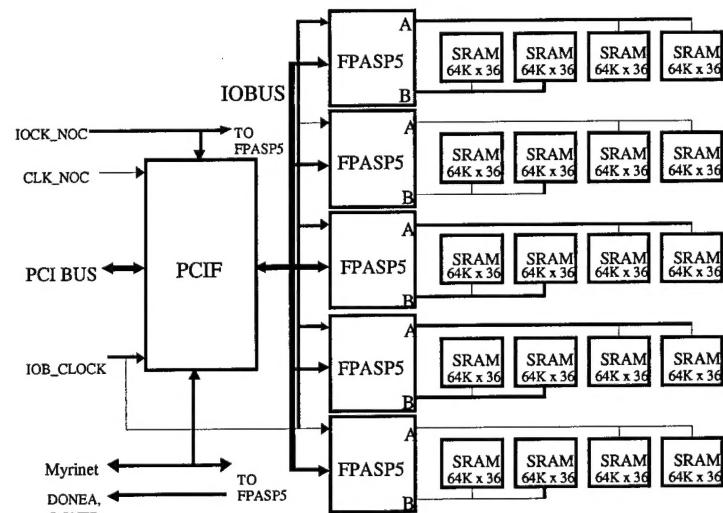


Figure 3. The IOBUS Configuration

clocks inhibited, the IOBUS has exclusive access to the element's resources. This is essential to ensure predictable messaging performance because the IOBUS does not have to wait for the processors to service an interrupt or even complete the current instruction, an important capability since a single vector instruction may require many cycles.

External Interface to the MCM

The external interface to the MCM can be implemented as a field programmable gate array (FPGA) or ASIC depending on speed and logic requirements. This point in the architecture is a natural place to incorporate local DRAM as shown in Figure 4.

UNCLASSIFIED

A set of eight state-of-the-art DRAMs can be attached to a controller within the MCM interface for a reasonable percentage of the overall MCM area. This overhead can be further reduced by using two short stacks of DRAM in lieu of eight discrete components.

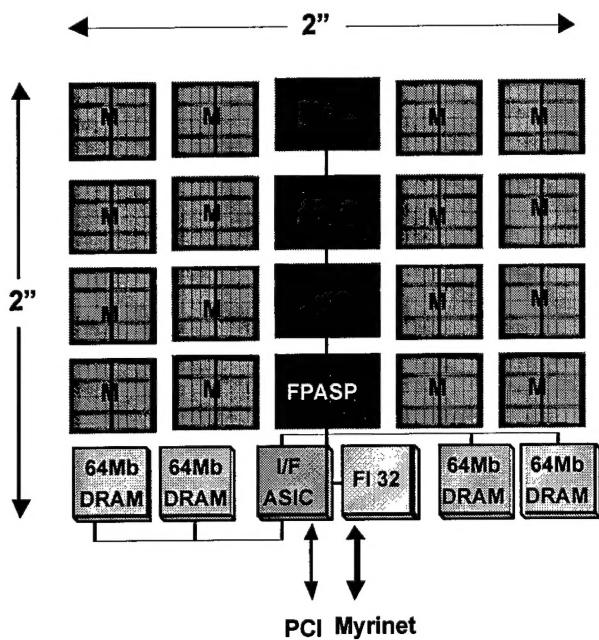


Figure 4. External Interface to the MCM

The primary purpose of the interface is to provide connection between the intra-MCM IOBUS and external standard buses. The 64 bit PCI bus is the natural choice due to its similarity to the IOBUS, and the availability of commercial hardware to bridge the PCI bus to many other high performance communications formats (e.g. Fibrechannel, Futurebus+, serial HiPPI, etc.).

High Density Interconnect (HDI) Packaging

The area HDI allows signals on the interior of the chip to be brought out to the HDI where it is convenient to do so. The photomicrographs of the WSSP processor and the ASIC interface (in Figure 5) show columns of pads going down the center of the die.

Using only pads around the periphery of the chip would require far more routing on the chip and its associated capacitance.

The area HDI removes the constraint that the pads be at the periphery of the chip because wire bonds are not used. This is particularly important in routing power. With the power available only

at the periphery, the voltage at the center of the chip may droop during periods of high current demand because of the inductance and resistance of the power network. Area HDI allows the power signals to distribute throughout the chip, and the copper metalization used in HDI routing is many times thicker than the aluminum metalization commonly used in VLSI. Because the HDI routing is thicker and may be much wider without sacrificing chip area, it provides lower impedance power distribution.

Area HDI also permits many more I/O signals to be taken off of the chip because the number of I/Os is limited by the number of HDI traces that can cross the boundary of the chip on a single layer of routing times the number of routing layers available. In conventional wire bonding, the number of I/O signals is limited by how many pads can be fit around the periphery of the chip. Since the trace pitch is typically less than the wire bond pad pitch, and the number of HDI routing layers is greater than one, area HDI can allow for far greater I/O signals.

In addition to reducing the amount of board area required for processors and their memories, HDI can reduce the overall power dissipation. Closer die-to-die spacing and fine lithography reduce routing capacitance and the power required to drive it. The lower power, in turn, may allow more chips to be packaged in one MCM because the power dissipation from driving inter-die signals may be significant.

One draw back of relying upon area interconnect is that the bare die are difficult to test with

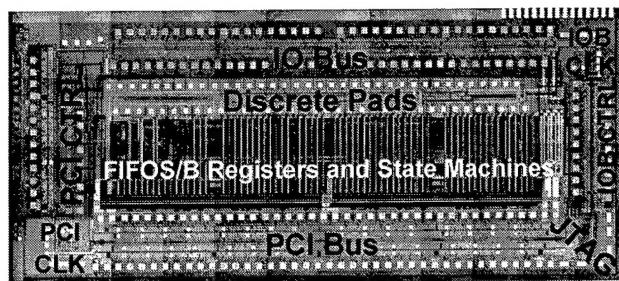


Figure 5. ASIC Interface Chip

wafer probes. Therefore, the WSSP testing is performed using Joint Test Action Group (JTAG) [6] boundary scan and internal scan testing.

Software environment

The WSSP processor is a fully programmable

digital signal processor. Its highly optimized vector routines complement a general-purpose RISC-like assembly language. The GNU tool set including the C compiler, assembler, linker, libraries, and debugger has been ported to the WSSP. A graphical debugger built upon the GNU is also available. To access the high performance vector instructions, Basic Linear Algebra Subroutines (BLAS) have been supplemented with additional common vector operations and encapsulated in C callable library functions for single and double precision arithmetic.

The extensible assembly language is defined via the VLIW control word, which gives fine-grained control of all resources on the processor, and allows high sustained performance on optimized routines. In addition, on-going research is exploring direct compilation from C into the writable VLIW control store. This eliminates the inefficiency of both the library call overhead and the assembly language fetch-decode-execute overhead.

The operating system is the Real-time Executive for Military Systems (RTEMS) [7], which provides an executive for real-time preemptive multitasking operating system. RTEMS supports multiprocessor systems with an event-driven or priority-based preemptive scheduler. It also supports intertask communication and synchronization. Work is currently under way to implement the MPI message passing standard [8] on the WSSP under RTEMS.

Application Examples

In signal processing, as in other fields requiring high performance computing, there are a few key operations that tend to dominate the overall throughput and latency computational requirements. The sustained performance of processors, such as the WSSP, as a percentage of peak performance on these operations provides useful benchmark information.

Figure 6 shows the WSSP processing element performance on a block update to a QR matrix factorization. This operation is typically used to update the model of clutter and interference in multi-channel sensors based upon newly received measurements. The horizontal axis is the number of new measurements (M). The number of channels (N) is fixed at 32. The complexity of the update is $O(N^2m)$. In a recent airborne radar experiment, M=50 new measurements were incorporated on each update. That application called for 128 such

updates to be performed as part of the overall signal processing comprising approximately 60% of the entire signal processing load [9]. The lower curve shows the performance for a C subroutine implementation calling the optimized library functions for vector operations. The upper curve shows the improvement when the entire block update is encapsulated in an assembly language library routine eliminating calling overhead and optimizing register utilization. For M=50, the C sustains 57% of peak and the library routine sustains 77% of peak.

Another important signal processing primitive is the Fast Fourier Transform (FFT), widely used to transform signals back and forth between the time and frequency domains. Unlike the matrix factorization, which has a balanced requirement for multiplies and adds, the FFT requires roughly twice as many adds as multiplies (6 vs. 4 for a radix-2 primitive operation, 22 vs. 12 for a radix-4 primitive). This leads to partial multiplier utilization in the WSSP, but still approximately 75% of peak utilization.

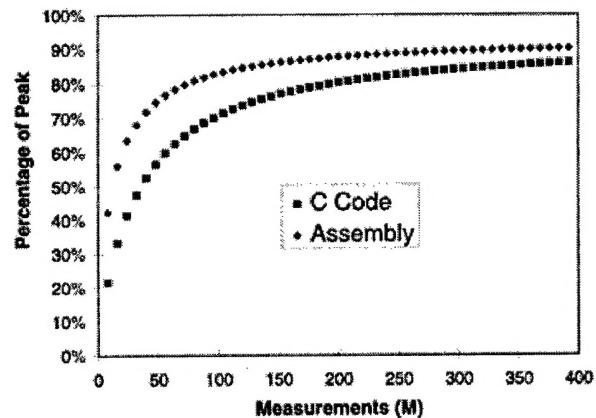


Figure 6. QR Matrix Factorization Efficiency

As these examples begin to show, the VLIW instruction word combined with the on-chip computational, address generation, and control resources are able to deliver a high percentage of the peak eight floating point operations per clock on important signal processing applications.

WSSP Configuration for DITP

The baseline configuration of the WSSP for the DITP application consists of eight MCM layers interconnected with a crossbar switch and a shared PCI bus as shown in Figure 7. Different MCM layers varying the number of elements and the

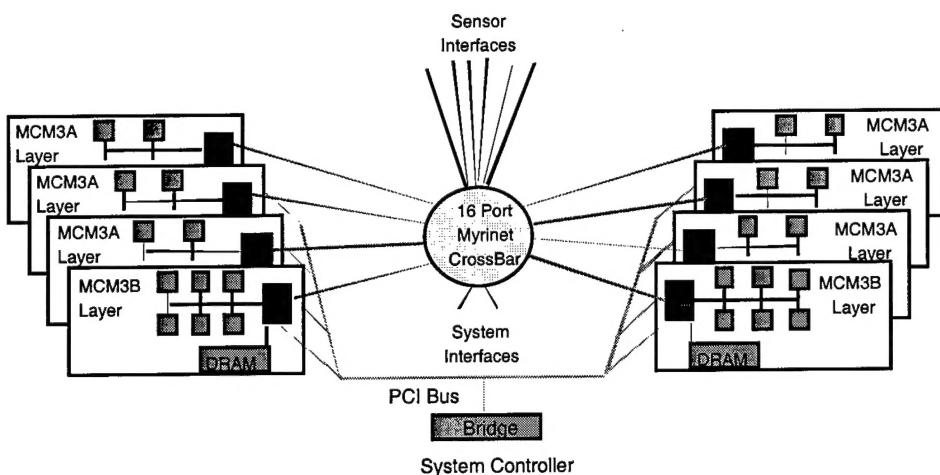


Figure 7. WSSP Baseline Configuration for DITP

amount of memory per element may be employed to match DITP requirements. At an anticipated 50 MHz clock rate on the WSSP elements, the 36 elements in the configuration shown provide a peak throughput of 14.4 GFLOPS.

MALLEABLE SIGNAL PROCESSOR

While high-performance signal processors are optimized for high-resolution integer and floating point operations, they are often ill-equipped to accommodate much simpler operations with comparable efficiency. Trivial operations frequently require similar intervals in time for such processors as do much more complex operations, such as a floating-point multiply. The operation and data handling of complex sensor resources, such as focal plane arrays characteristically require multiple operations of relatively simple complexity but high concurrency. While in principle processors such as a WSSP could accommodate such processing, these computations would consume a number of such powerful processors with simplistic logic and sequencing operations, blocking access to most of the useable silicon due to the cycle-oriented nature of stored program execution in von Neuman processors.

The MSP was conceived of to provide the ability to support digital logic interfaces between a variety of complex sensors and more powerful general purpose signal processors. The MSP (shown in Figure 8) employs a reconfigurable processing approach, in which the logic configurations of a number of RAM-based

FPGAs are defined at run-time. This run-time configuration is done through the use of a configuration management function, which establishes the "personality" of the MSP based on the type of sensor at the left of the MSP. Since the logic configurations in MSP can in principle be changed as a function of which sensor and fusion processor interface are chosen, this intervening block of functionality is malleable, hence the name.

Groups of independent, simple state machines and combinational logic are readily formed, achieving the necessary concurrency to implement a variety of simple time-intensive logic functions.

Design / Architecture

The present MSP design consists of: (1) a sensor adaption section; (2) a MSP core; (3) a high-speed Myrinet interface; and (4) a configuration management processor.

Sensor adaption section. In the present MSP concept, true "malleability" is endowed only to CMOS-level digital functions and signals (bi-level voltages of 0V and 3.3V). For a complex imaging sensor, particularly a cryogenic FPA, most I/O, even bi-level ones, rarely meet this requirement. Often, signals in cryogenic FPAs are slew-limited, and their bi-level representations are chosen for convenience in implementing the read-out integrated circuit (ROIC) of the FPA. As such,

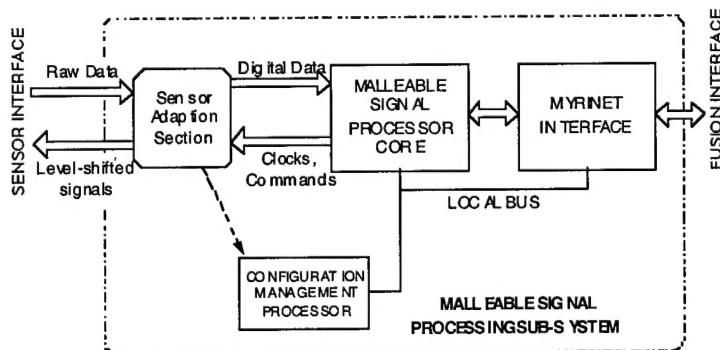


Figure 8. Malleable Signal Processor (MSP) Concept Diagram

UNCLASSIFIED

rather than having a (0V, 3.3V) bi-level, compatible with an MSP, a particular FPA may have one set of signals at one bi-level (-3V,+2V), another set of signals at a second bi-level (+3V,+4V), and a third set of signals at yet another bi-level (-3V,+3V). Furthermore, some signals may be differential bi-levels for improved noise immunity. Finally, many cryogenic sensors will have multiple analog outputs, which will require digitization in order for a MSP to deal with them.

The *sensor adaption section* of the MSP represents the only section particular to a given MSP interface. The types of signals implemented by a sensor adaption section can be broken into three types: (1) clock and command signals, (2) bias signals, and (3) variable analog signals [10]. Clock and command signals are defined as static or time-varying bilevel signals that by definition must be "laundered" to the (0V, 3.3V) CMOS-compatible bi-level that the MSP can accept or generate. Bias signals are defined as precision direct current signals used to provide power and reference voltages to a FPA. FPAs require from 1 to 24 bias signals, each with potentially a different value. Finally, most FPAs produce analog signals, which require digitization. Typical FPAs have from 1 to 16 outputs, correspond to distinct temporal or spatial subsets of a large pixel (x, y) field, available in one or more response wavebands.

MSP Core. The MSP core consists of a large block of RAM-based FPGA logic and data memory. The core of the present MSP design (version 0.1) is shown in Figure 9.

The MSP is arranged as a nearly symmetric pipeline, with the intention in general application of flowing data from the sensor (left) to a fusion processor (right). The 200,000 gates of reconfigurable processing are based on two Altera 10K100A devices, which are each configured over a customized local bus upon power-up. Each Altera device controls 12 megabits SRAM using two independently addressable banks organized as 512K x 12 bits of storage. The SRAM banks are designed to retain data, if necessary, during reconfiguration of one or both FPGAs, permitting in-situ changes of personality during run-time.

Although the in-system reconfiguration of entire FPGAs is possible, the process requires at 100 ms, and is more intended to conform to a static reconfigurable processing model than a dynamic context-switch model. The ability to reconfigure during operation is

nevertheless very important, as it allows for diagnostic, calibration, and coefficient loading during operation, where the rotation between such modes can be considered virtually static relative to the timelines of in-flight operation.

Myrinet Interface. The high-speed Myrinet interface, as a de facto standard in DITP, is used to pass high speed data streams from the MSP to the fusion processing network, as well as relatively low-bandwidth commands in the reverse direction. As such, a bridge between MSP and Myrinet, a high-speed (~gigabit/sec) low voltage differential signal protocol, is required. The implementation under development for the MSP in DITP involves a Myricom FI32 component, which provides a direct translation of high speed CMOS data to the Myrinet protocol. The FI32 component is driven by the MSP core through a third Altera 10K100 device resident within the FI32-based interface. This third 10K100 is completely dedicated to implementing a virtual first-in/first-out (FIFO) buffer and "normalizing" the bus structure generated from the MSP core for high speed data transport.

Configuration Management Processor. MSP initialization and "personality management" are handled through a low-level dedicated processor in the embedded DITP system. This configuration management processor contains: (1) a local bus interface for driving the local bus of the MSP core, (2) a large store of non-volatile flash memory; (3) a sensor identification port; (4) clocking circuitry for the MSP core; (5) and two system interface serial ports. The flash memory stores all required personalities of the MSP for operation, along with any data tables (such as gain and offset coefficients for a FPA). The sensor identification port effects a very sophisticated "plug-and-play" mechanism, whereby the MSP can actually be

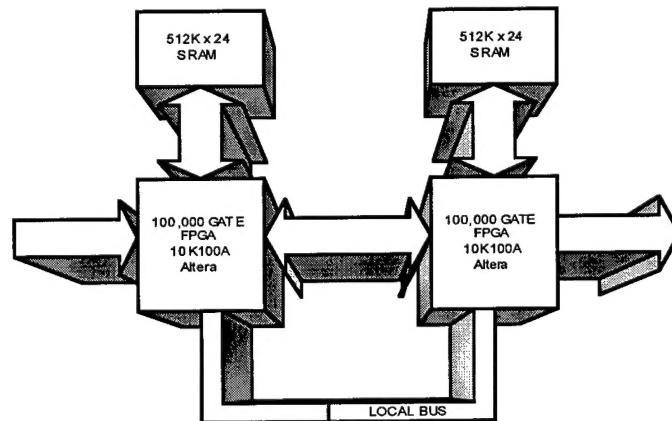


Figure 9. Core of the MSP 0 Design (version 0.1)

interfaced to several different sensors without changing the MSP's hardware or internal programming. This definition of "plug-and-play" does, however, require that the sensor adaption section be considered integral to the sensor assembly. In this manner, interfaces can be reconfigured at runtime by physically unplugging one sensor and plugging in another at the sensor adaption section-to-MSP core interface.

Operationally, the MSP would operate complex imaging sensors at high frame rates, providing non-uniformity corrected raw data to the fusion processor. At the time of this writing, a dual-waveband 256x256 FPA is baselined for flight, although a number of other candidates are possible for back-up without impacting the MSP core, MSP management (MMGT), or Myrinet interface. In-situ reprogramming of the MSP non-volatile store system is possible, which permits personalities to be refined, added, and deleted from embedded memory. The MMGT serial port operation is conceptually script-like, an extension of an embedded monitor approach. For flight, a default script can be executed after a time-out period has elapsed, during which an external monitor could over-ride the default program for test and maintenance. Calibration of the sensor is similarly interactive, aided by the ability of the MSP core to be programmed to capture FPA data at different background levels. This capture information can be processed by a host computer present during debug to generate gain and offset "images" that can be "locked" into MMGT firmware. The calibration process can be repeated as required, even immediately before launch. The MMGT operation is also reconfigurable *in situ*, adding yet another level of flexibility to the MSP system.

HIGHLY INTEGRATED PACKAGING AND PROCESSING

The packaging system for DITP is termed "Highly Integrated Packaging and Processing" (HIPP), which represents the world's most advanced heterogeneous packaging approach for complex systems. A long-standing relationship involving BMDO, and formerly separate organizations with the Air Force Research Laboratory has existed over a decade leading to the co-development of HDI technology, for which significant funding from DARPA and General Elec-

tric was also involved. This technology was recently applied successfully to over 50 designs from AFRL and BMDO, including the WSSP [11].

A painful lesson learned by many over the last decade is that MCMs alone provide only a partial solution to packaging advancement. Often, MCMs are invoked as "cure alls" in a technology development program, yet in many cases the MCMs do not achieve the dramatic improvements suggested by simply comparing the MCM substrate to the bulk of the packaged components replaced. The key to achieving system-level improvements in electronics density clearly lies at the higher levels of what is referred to as the "packaging hierarchy". Printed wiring boards, boxes, and even the system platform itself interact with and effect the density of packaging possible.

Traditional vs. HIPP Packaging Hierarchy

The traditional packaging hierarchy has occurred through "happen-stance" over the last four decades (an example shown in Figure 10). Its present form is convenient and represents the embodiment of most modern electronics. As will be discussed, however, the traditional hierarchy is limiting at higher levels (above level 2), representing the substantial reliance on monolithic integrated circuits for rendering the vast majority of today's electronics [12].

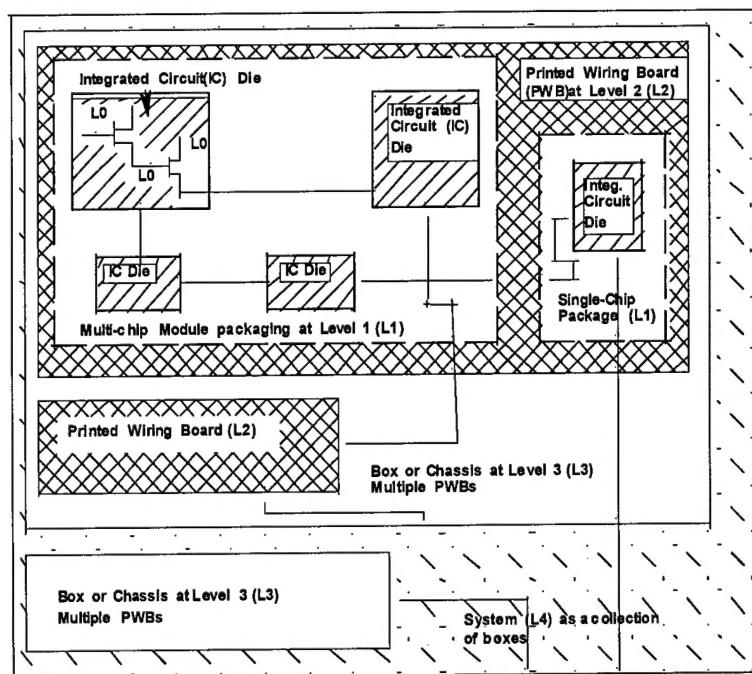


Figure 10. Traditional Packaging Hierarchy

In the traditional packaging hierarchy, level 0 refers to interconnections between individual transistors on an integrated circuit. Level 1 interconnect refers to the transition from diced wafers to package. Level 2 interconnect refers to the package-to-package or printed wiring board (PWB) interconnect. Level 3 interconnect refers to commonly the inter-chassis board-to-board interconnect. Finally, level 4 refers to the system platform when viewed as a collection of electronics chassis components linked together by cables and connectors.

DITP HIPP Assembly Structure

The HIPP program has sought packaging solutions more optimal at a system level, based on the concept of closely integrating a collection of various MCM substrates or other assemblies of identical size and conductor arrangement. The requirements of candidate HIPP structures include: (a) the ability to accommodate high I/O densities (up to 1000 per MCM layer); (b) support of heterogeneous signal composition (i.e., analog signals alongside digital, power, and microwave signals); (c) modularity and service-ability for layer repair and replacement; (d) adequate power and thermal management; (e) adequate I/O density at the second level of packaging; and (f) robustness for applications in harsh environments. Continued research at AFRL under joint BMDO fund led to the development of a candidate heterogeneous 3-D packaging system. Such a system, the baseline for DITP, is shown in Figure 11.

Figure 11 illustrates a many-layer 3-D packaging approach that combines a number of *segment* entities into an *assembly*. The segments, which are the common and fundamental building block of HIPP, contain one or more MCMs or small circuit boards containing components. Systems, such as the DITP interceptor platform, can be partitioned into a number of segments, as shown in Figure 11a.

In this case, the HIPP assembly baselined for DITP consists of approximately 16 segments

(numbers reflect present order of segments from the front in the preliminary design):

- MSP Subsystem 1: Sensor Adaption segment for passive focal plane array sensor (#2); MSP 0.5 core segment (#4); MMGT segment, which contains MSP management processor and FI32 interface (#5)
- MSP Subsystem 2: Ladar Adaption segment containing non-digital interface (level shifters, etc.) (#3); MSP 0.5 core segment (#6); MMGT segment for second MSP core (#7)
- Fusion Processing Subsystem: Myrinet Crossbar (#10); Wafer Scale Signal Processor Segments (WPSG) (#10 and #11); Wafer Scale Signal Processor Management Processor (WMGT) (#13)
- System / Misc: Front (#1) and back (#16) connector segments; Power management segments (#13) for fusion and (#9) for MSP subsystems and servo/guidance interface; Servo layer, which contains interfaces to

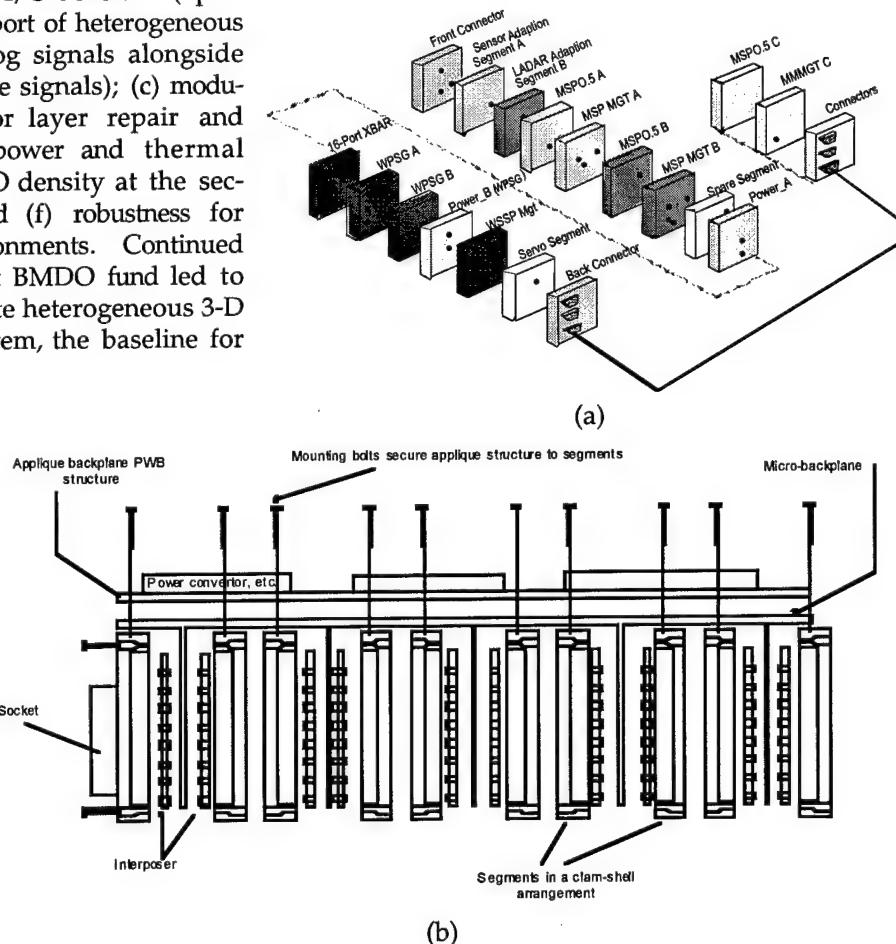


Figure 11. HIPPI Baseline Concept for DITP:
 (a) Simplified Physical Representation;
 (b) More Detailed Drawing Illustrating Key Features

communicate with servos, guidance (#15); Spare segment (#8)

These 16 segments, referred to collectively as the Sensor and Fusion Engine (SAFE), have a common substrate size and I/O pad location. The physical size for substrates used in the DITP version of HIPP is 2 x 2 inches, and up to 1000 I/O can be accommodated on the surface of each segment.

The contents of each segment can be completely different, and in fact the type of MCM technology used in each segment can be different so long as the segment definition is not violated. As such, layers do not necessarily need to be based on MCMs, but in fact could be single-chip packages, small printed wiring boards, hybrids or MCMs. In the terminology of HIPP, segments are said to contain one or more *layers*. Examples of possible layer arrangements are shown in Figure 12.

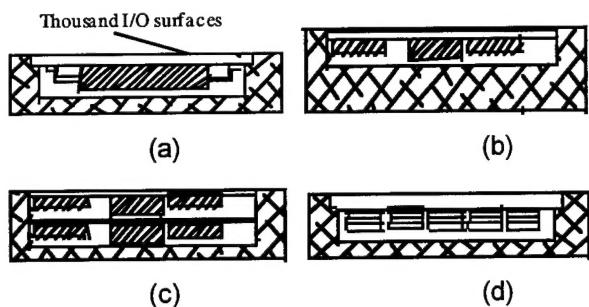


Figure 12. Generic HIPP Segments.

ments within segments are shown in Figure 12. Figure 12a illustrates a single-layer segment containing a single component mounted on a printed wiring board. Figure 12b illustrates a single-layer segment containing one MCM. Figure 12c is an example of a segment with multiple layers, in this case two high density interconnect MCMs. In principle, extremely dense MCMs could be used in multi-layer arrangements as shown here to increase the volumetric densities of individual segments over that possible with a single-layer segment. Finally, in Figure 12d is shown a single-layer segment in which a number of densely stacked 3-D chip configurations have been placed. In this manner, the HIPP packaging technology is versatile in that it can accommodate many existing modern forms of single-chip, multi-chip, and 3-D packaging.

The Micro-backplane and Assembly

The deliberate stacking of segments is referred to as an *assembly*. Figure 11b illustrates some of

the special structures needed to integrate multiple segments into a complete assembly. These structures include: the micro-backplane, a number of interposers, an applique superstructure option for attaching more connectors or electronics, and hardware for secure segments to the micro-backplane. The micro-backplane that uniquely defines the pattern of all inter-connections in the DITP SAFE system. The micro-backplane is a compound flex system, based on a long manifold of multi-layer copper-polyimide with orthogonal tabs of flex that address the face of every segment in the HIPP system. The micro-backplane can be thought of as the "nervous system" of a HIPP assembly. Figure 11b illustrates the notion of clamshell mounting, a technique by which particular segments are mounted face-to-face through the micro-backplane. Clamshell mounting allows a more intimate interconnection between two particular segments, which can serve to reduce complexity in the micro-backplane.

The interposers provide a compliant contact system, such as required when a large number of contacts between two flat surfaces must be brought together. Since HIPP segments may contain up to 1000 I/O, then so must the interposers. JEDEC standards for ball grid array are being exploited for HIPP segment and interposer I/O definitions. The concepts of segments, interposers, and micro-backplanes are shown in Figure 13.

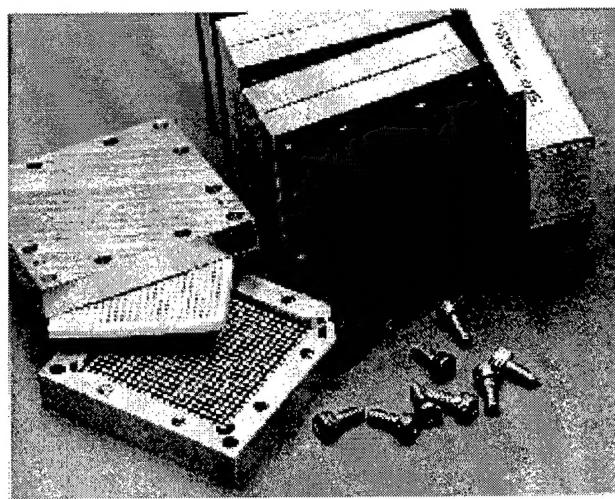


Figure 13. Experimental HIPP Assembly.

Power and Thermal Management

As a heterogeneous packaging system, the HIPP approach must deal with extremes in power management, thermal management, and signal

integrity. Power delivery concepts under definition for HIPP are addressing the problem of delivery of 70-80 amperes of current at 3.3V, most of which is consumed by fusion processing. The power supplies for systems with a large content of switching digital face a significant challenge in combating simultaneous switching noise. Also, the delivery of large amounts of current could require heavier metal structures than those available in a micro-backplane, and I^2R losses on 3.3V systems have more impact than on 28V delivery. Efficiency in power conversion is a primary driver in the fusion processing hardware. On the other hand, small amounts of very stable current are required for analog sensors. Efficiency in this case is much less important than stability. As such, power management and distribution requires careful consideration. In the DITP system, at least two separate segments are devoted to power distribution and management, one for fusion processing and one for all other electronics.

The thermal management philosophy in HIPP is hierarchical, based on first shuttling heat generated within each segment efficiently as possible to the outer edges of the segment walls and then coupling a second-level thermal management system.

At the first level, the burden of thermal management responsibility lies with the segment designer. In the SAFE system, segment power dissipation range from about 1.5W to 60W. Thermal transport in HIPP must occur laterally, parallel to the plane of the layers within the segments. This is because many HIPP assemblies, such as the DITP SAFE are *many-layer* MCM assemblies, in which more lateral surface area (in this case the segment walls) is available for heat conduction than the top or bottom of particular substrates. Figure 11 illustrates the segment level thermal path. In HIPP assemblies, a maximum of three out of four segment walls can be devoted to thermal management. Segment thickness, segment wall thickness, and segment material selection can be based on local and global HIPP assembly needs. It is also possible to consider more exotic, active thermal management approaches within segments to improve thermal transport at the segment level.

The second level of thermal management is application dependent, but must deal with power

dissipation levels as high as 500W for 16 segments in the highest performance designs (the DITP safe dissipates about 200W). In the DITP SAFE assembly, as suggested in Figure 14, lateral heat transport from segments into a phase change material is a second-level thermal management approach under consideration. For short-term missions, such approaches can be considered. For operation at longer intervals, a number of other options can be considered, ranging from heat sinks to heat pipes and liquid flow-through systems. Thermal management systems can more intimately link into segment walls, through texturing, flocking, insertion of flow through channels, etc. Such concepts are helpful in reducing thermal transport between segments and the secondary thermal management system.

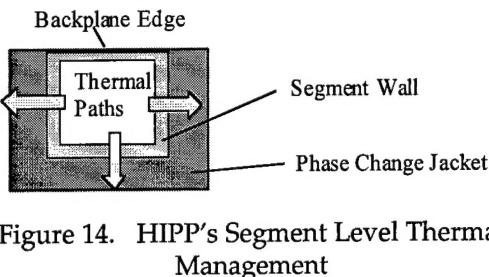


Figure 14. HIPP's Segment Level Thermal Management

A New Packaging Hierarchy and Extensions of the HIPP Framework

HIPP establishes an alternate packaging hierarchy, compatible with and more efficient than the existing hierarchy. Level 1 in the HIPP packaging hierarchy refers to internal layer composition, level 2 is defined by layers within segments, level 3 is defined by the segment itself, and level 4 is defined as the HIPP assembly of segments. Various forms of compatibility with the existing packaging hierarchy are readily achieved. For example, HIPP segments can be face-mounted onto printed wiring boards through proper socketing or through conversion of the land grid array on the segment face to a ball grid array. Alternately, entire HIPP assemblies can be mounted onto a printed wiring board, given the proper structural design. HIPP assemblies can be used to replace entire boxes; miniature connectors can be introduced on the front and back surfaces and onto the micro-backplane itself.

HIPP offers a packaging system which meets the essential requirements of a 3-D modular packaging system with high efficiency and flexibility. Segments are in essence inter-changeable, repairable, and replaceable without great difficulty. By employing designs for segment I/O patterns that are compatible with JEDEC definitions, HIPP can leverage the substantial investment in ball grid array (BGA) technology, which has a reasonably well-established infrastructure for fixturing that

UNCLASSIFIED

can be exploited in the development of breadboard test systems. In contrast to many 3-D packaging approaches, HIPP can inter-mingle a great diversity of functional domains in electronics, providing in this way a great flexibility for system designers. Design guidelines under development will in time reduce to practice many aspects of the analysis needed to effectively use HIPP technology, such as electrical and thermal design rules and guidelines for exploiting computer-aided-design (CAD) tools. It is believed that with these guidelines it will be possible to design HIPP-based systems with the ease and confidence one would use to design present-day VME or SEM-E board-based systems.

SUMMARY

For the DITP application, a total sensor fusion processor system will consist of an assembly of 12 to 15 such segments that are butted together with clamping rods. A dimension of the integrated processor package is approximately 2 inches square on the sides by 4 inches long. The HIPP is being given a stressing test — combining the necessary processing power for an advanced on-board fusion of data in real-time from several complex imaging sensors and a LIDAR. This requires developing a 2" x 2" x 4" microsystem containing most electronics of an end-to-end seeker, including sensor (dewar) interfaces, 14 to 20 GFLOPS of processing, servo drive, and system management. The HIPP system will have a thermal design capacity of 500 W (although the actual architecture is expected to dissipate much less power), a maximum interconnect capacity of 30,000 nets between all layers, and will represent at least an order of magnitude or more density improvement over any previous demonstration of its class. The DITP application of HIPP will constitute the densest heterogeneous 3-D system assembly attempted to date.

The HIPP system, as an MCM-containment framework, provides opportunities for further density enhancement by improving the constituent MCM technologies which are placed within it. One such approach is the Ultra-High Density Interconnect (UHDI) technology, which provides an ultra-thin MCM capability. Combining UHDI with HIPP results in a potential to further quadruple the density of the architecture. UHDI is being investigated within the DITP architecture and other applications in creating extremely dense internal memory storage for the WSSP layers of the architecture. Also, the HIPP and WSSP tech-

nologies were originally designed for the BMDO to operate in a harsh space radiation environment. There is a potential plan for a parallel development of the radiation-hardened version of the processor being contemplated by the BMDO and AFRL.

In Summary, the DITP's Sensor and Fusion Engine (SAFE), consisting of the WSSP/MSP architecture with the HIPP integration, is a major advancement toward a flexible (and reconfigurable) massively-parallel floating-point MIMD architecture with an innovative 3-D heterogeneous miniaturized packaging to meet the most demanding and diverse multi-functional processing requirements for the future system applications.

ACKNOWLEDGMENT

The authors wish to acknowledge the sponsorship and direction provided by LCDR James (Chip) Buckley, BMDO/TOS, and technical guidance provided by Dr. Walter Dyer, BMDO/TO. Also much appreciation is extended to Mr. Bill Fontana, AFRL/IFTC, for ably filling his role as an executing agent (EA) for managing the DITP Fusion Processor and Algorithm development task under the DITP. This applies equally to Dr. Paul LeVan (AFRL/VSSS) for his role as an EA on Passive Sensor Subsystem (PSS) development effort.

REFERENCES

1. Y. Kinashi, R. W. Linderman et. al., "Discriminating Interceptor Technology Program (DITP): Sensor Fusion for Improved Interceptor Seekers," 1996 AIAA/BMDO Missile Sciences Conference, Dec. 1996.
2. R. A. Fillion et. al., "Development of a Plastic Encapsulated Multi-chip Technology for High Volume, Low Cost Commercial Electronics", IEEE Transactions on Components, Packaging, and Manufacturing Technology Part B: Advanced Packaging, Feb. 1995, Vol. 18, No. 1, pp 59-65.
3. PCI Special Interest Group, "PCI Local Bus Specification, Revision 2.1", PCI Special Interest Group, Portland, OR, 1995
4. G. Blalock, "Microprocessors Outperform DSPs 2:1", Microprocessor Report, *Micro Design Resources*, pp 12-15, Dec. 30, 1996.
5. ANSI/IEEE Std 896.1-1994, FutureBus+ Logical Protocol Specification.

UNCLASSIFIED

6. C. M. Maunder, R. E. Tulloss, "The Test Access Port and Boundary-Scan Architecture", *IEEE Computer Society Press*, 1990.
7. RTEMS Homepage, <http://lancelot.gcs.redstone.army.mil/rtems.html>.
8. M. Snir, S. W. Otto, S. Huss-Lederman, D. W. Walker, J. Dongarra, MPI: The Complete Reference, The MIT Press, Cambridge, MA, 1996.
9. M. H. Linderman and R. W. Linderman, "STAP Demonstration on an Embedded High Performance Computer", to appear in 1997 IEEE National Radar Conf., May 1997.
10. Private conversation with Mark Stegal, SEIR Corporation, April 1998.
11. R. W. Linderman et. al., "HDI Design of A 100 MFLOP/Watt Floating Point DSP," Proceedings of Government Microcircuits Application Conference, March 1998.
12. J. C. Lyke, "Packaging Technologies for Space-based Microsystems and Their Elements," Chapter in Microengineering Technology for Space Systems, Aerospace Report No. ATR-95(8168)-2, 30 Sep 95, edited by Henry Helvajian.